DOCUMENT RESUME

ED 221 579 TM 820 608

AUTHOR Gabriel, Roy M.
TITLE Statewide Implementation of the Value Added Model: Methodological Problems and Prospects.
PUB DATE Mar 82
NOTE 50p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Development; *Data Analysis; *Early Childhood Education; Models; *Program Evaluation; *Research Methodology; *Research Problems; Statewide Planning
IDENTIFIERS Aggregation (Data); Evaluation Problems; Oregon; RMC Models; Sample Size; *Title I Evaluation and Reporting System; *Value Added Model

ABSTRACT

Data from 86 Early Childhood Education (ECE) projects in Oregon were analyzed using methods prescribed by the Value Added model for ECE evaluation. The extent to which appropriate model implementation procedures were followed is unknown. Thus, the results represent an empirical account of the model's estimates of project impact, rather than a rigorous test of the validity of the model. In these data, estimates of student's cognitive growth rates at the project level were highly unstable, primarily due to a preponderance of small sample sizes. Methods of aggregating data across projects were examined, with resultant estimates gaining stability, but still seemingly low. Comparisons of the Value Added model's estimate with that of the Title I Evaluation and Reporting System Norm Referenced model supported this contention. Other issues explored in these data were standardizing the outcome metric for large scale reporting purposes, the utility of additional student background information on the model's estimate of growth rate and the characteristics of the variety of 18 tests represented, whose content, administration and scoring influenced the applicability of the model. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED221579

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Statewide Implementation of the Value Added Model: Methodological Problems and Prospects

Roy M. Gabriel

Title I Technical Assistance Center
Northwest Regional Educational Laboratory

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. M. Gabriel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 608

Paper presented at the annual meeting of the American Educational Research Association, March 19-23, 1982. New York, New York.

ACKNOWLEDGMENTS

The author is indebted to a number of colleagues for their assistance and support in bringing the work represented here to its current status. Initially, Joy Martin bore much of the responsibility for creating, screening and editing the data base used in this study. Given the lack of involvement in the early stages of planning the data collection activities, hers was an arduous and often frustrating task. Connie Fitzgerald worked through hordes of project and aggregate-level impact calculations. Her observations and insights during these otherwise routine tasks added significantly to the interpretation of results presented here, and provided valuable insights for further exploration of issues in these data. The collaborative efforts of Dr. Sharon Harpring, a co-investigator in this project and author of a companion paper, consistently brought clarity and coherence to the interpretation of these results from a different experiential perspective than that of this author. Her familiarity with early childhood education students and classrooms is an invaluable asset in this and future methodological studies in ECE evaluation.

ABSTRACT

Data from 86 Early Childhood Education (ECE) projects in Oregon were analyzed using methods prescribed bt the Value Added model for ECE evaluation. The extent to which appropriate model implementation procedures were followed is unknown. Thus, the results represent an empirical account of the model's estimates of project impact, rather than a rigorous test of the validity of the model. In these data, estimates of student's cognitive growth rates at the project level were highly unstable, primarily due to a preponderance of small sample sizes. Methods of aggregating data across projects were examined, with resultant estimates gaining stability, but still seemingly low. Comparisons of the Value Added model's estimate with that of the TIERS Norm Referenced model supported this contention. Other issues explored in these data were standardizing the outcome metric for large scale reporting purposes, the utility of additional student background information on the model's estimate of growth rate and the characteristics of the variety of tests represented which influenced the applicability of the model.

Evaluating the impact of early childhood education (ECE) programs invokes a collection of methodological issues not common to those involved in the same task applied to other educational programs. Evaluation models developed for ESEA Title I programs in the basic skills, grades 2-12 (Tallmadge & Wood, 1976), rest on assumptions and involve procedures which are often not reasonable with school (and pre school) children of a younger age. The assessment of school achievement, for example, presents a different challenge, due to the special problems inherent in testing very young children (Haney & Gelber, 1980). The stability of testable cognitive characteristics is of greater question, and the normative data is of far less standard quality than those of standardized tests for elementary and secondary school children.

The Value Added Model has been proposed as one method for evaluating the effectiveness of ECE programs. It views an ECE program as a "dynamic intervention in an ongoing developmental process" (Bryk & Weisberg, 1976, p. 129) and attempts to model this developmental growth rate on the outcome criteria of interest to establish a no-treatment expectation. Post-treatment outcomes are then compared with these expectations to determine the impact of the program. In the terminology employed in the Title I Evaluation and Reporting System (TIERS), the impact of the program is represented as follows:

Program Impact = (Post-treatment outcome) - (No-treatment expectation)

The modeling of developmental growth is particularly critical in dealing with children at such a young age. It ranges from a simple bivariate regression of pre-treatment criterion measures on age to more

complex models involving background variables on the project students (e.g., sex, ethnic group, parental education) which may more fully account for the developmental influences on the criterion measures (Bryk, Strenio & Weisberg, 1980).

Unlike the evaluation models used in the Title I Evaluation and Reporting System, the Value Added strategy requires neither the use of a norm referenced test nor the presence of a no-treatment comparison group. The validity of its estimate of a no-treatment expectation hinges on the statistical modeling of developmental growth of project students. Bryk and Woods (1980) offer a lucid and thorough discussion of potential difficulties of effectively modeling these processes in actual practice.

The purpose of this paper is to present empirical results of the use of the Value Added model for evaluating the impact of ECE programs in a number of school districts in Oregon. Issues previewed by the model's developers are examined on these data, and a few new considerations suggested for its implementation and the interpretation of results.

Harpring (1982) provides a detailed overview of the initiation of these efforts in evaluating Oregon ECE programs, and comments on some of the major implementation-related concerns for the utility of the model's application. This paper will touch on these concerns only insofar as they affect the nature of the data reported, and its interpretation via the Value Added analysis.

DESCRIPTION OF SAMPLE DATA

Test results and relevant project and student background information from 86 ECE projects in 36 school districts across the state were included in these analyses. The majority of these programs were

classified (by local project directors and teachers) as reading programs, and most occurred at the first grade level. Over half of the students were males, with pre-program ages ranging from 45 to 105 months. Table 1 contains more specific descriptive statistics on the total sample of early childhood education students. Of note is the extent of missing data evident on the student variables. Fortunately, the purpose of this study was not to represent the population of Oregon ECE students or the statewide impact of Oregon ECE programs. Again, the purpose here is to test the feasibility of the model for estimating impact in a "real world" application of an ECE program. Of some relevance to this purpose is that, of the 1,068 students (a duplicated count, i.e., a student in both a reading and math program was counted twice) reported in the entire sample, 956 (90%) had data on students age and pretest score--the most basic ingredients for using the Value Added strategy for estimating a no-treatment expectation. Again, Harpring (1982) presents a more thorough discussion of the quality control efforts undertaken in the data collection process.

Insert Table 1 about here

A list of the tests employed by the 86 projects in the analysis sample is presented in Table 2. In all, eighteen are represented, one of which is a locally developed inventory of developmental skills (Early In, 1976). The variety of tests and their relevant characteristics (content emphasis, administration procedures, scoring directions) posed a number

Table 1

Descriptive Statistics of
Oregon ECE Student Sample

	N	%
Sex		
Male	576	53.9
Female	380	35.6
Missing	112	10.4
Grade		
Pre-K	46	4.3
K	318	29.8
1	704	65.9
Missing	0	0.0
Subject Matter		
Reading	729	68.3
Math	148	13.9
"Other"	191	17.9
Missing	0	0.0
Preschool Experience		
Yes	71	6.8
No	246	23.6
Missing	726	69.6

Pretest Age

Minimum = 45 months
Maximum = 105
Median = 74.25
Mean = 73.13
Std. Deviation = 9.38

of interesting issues in the large scale implementation of the Value Added model discussed here.

Insert Table 2 about here

Preliminary correlational analyses of the 86 projects revealed a serious, though not unanticipated, problem in applying the Value Added model to these data. Figure 1 contains a frequency distribution of the project-level correlations between age and pretest scores. They literally span the continuum from -1 to +1, with approximately 40% being negative. Since the regression coefficient, relating age to pretest and providing the model's estimate of growth rate, agrees in sign with the correlation, we are left with the patently unsatisfactory situation of a negative growth rate estimate. In other words, our prediction--in these 40% of the cases--would be that project students would decrease in the skills measured by the pretest during the course of the program. Any of a number of characteristics of the sample (e.g., outlying values: very high/low age coupled with very low/high pretest scores) or instrument may be brought to bear in explaining these unlikely estimates. A very simple one is suggested in these data, however--a preponderance of small sample sizes, well known to be the bane of parameter estimation. Figure 1 also contains an indication of the average sample size in the projects exhibiting correlations in each of the intervals. Smaller sample sizes are visibly concentrated in the more extreme values of age by pretest correlations. More specifically, all correlations less than zero are associated with average sample sizes of 10 or less.

Table 2

Frequency of Tests Used in
Oregon ECE Project Sample

Test	# of Proj.	# of Students	% of Students
MAT(S) '78	25	262	25
Gates '78	14	171	16
MRT '76	9	133	12
Peabody PVT '59	6	119	11
Early In '76	4	51	5
PIAT '70	4	41	4
SRA '78	4	27	3
Boehm '71	3	55	5
Santa Clara '74	3	52	5
CTBS '73	3	16	2
CAT '77	3	14	2
MAT (I) '78	2	18	2
Inv. Readiness Skills '73	1	78	7
MRT '65	1	12	1
Woodcock '73	1	8	1
MAT '70	1	8	1
Gates '65	1	3	1

Insert Figure 1 about here

To further illustrate the influence of sample size on the estimate of growth rate, a scatterplot of these two values across all projects is presented in Figure 2. The range of these estimates is shown to be greatest for small sample sizes, monotonically narrowing as N approaches 10, with no negative estimates for samples greater than 16.

Insert Figure 2 about here

Aggregation Methods

With samples of this size, and growth rate estimates of such instability, calculations of project impact using the Value Added model are highly suspect. Since a high proportion of Title I projects across the country share this characteristic, it is an issue of major import in using the model. We chose to consider ways in which data from a number of projects could be pooled to obtain more stable estimates of growth rate and, subsequently, project impact. Thus, the prospect of aggregating data across projects--no stranger to the Title I Evaluation and Reporting System itself--became a central issue in this study.

The aggregation schemes employed here are of a slightly different character than those of the TIERS. In the latter, estimates of project impact obtained from the evaluation models are expressed in a common reporting metric (Normal Curve Equivalent scores) and are aggregated across projects, and eventually across states, to obtain a state, and national, estimate of the impact of Title I projects. In other words,

Figure 1

Distribution of Project-Level
Age/Pretest Correlations (r)

Interval of r value	f_i	$\bar{N}$
Greater than .91	4	3.5
.81 to .90	0	-
.71 to .80	3	12.7
.61 to .70	2	9.5
.51 to .60	3	12.3
.41 to .50	3	16.3
.31 to .40	7	12.8
.21 to .30	8	14.8
.11 to .20	9	28.2
.01 to .10	5	16.4
.00 to -.10	4	10.5
-.11 to -.20	6	10.2
-.21 to -.30	1	6.0
-.31 to -.40	4	7.5
-.41 to -.50	5	5.8
-.51 to -.60	0	-
-.61 to -.70	1	7.0
-.71 to -.80	3	5.7
-.81 to -.90	3	4.7
-.91 to -1.00	4	2.8

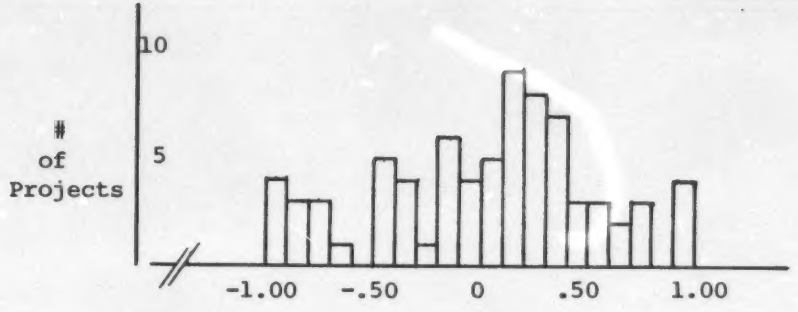
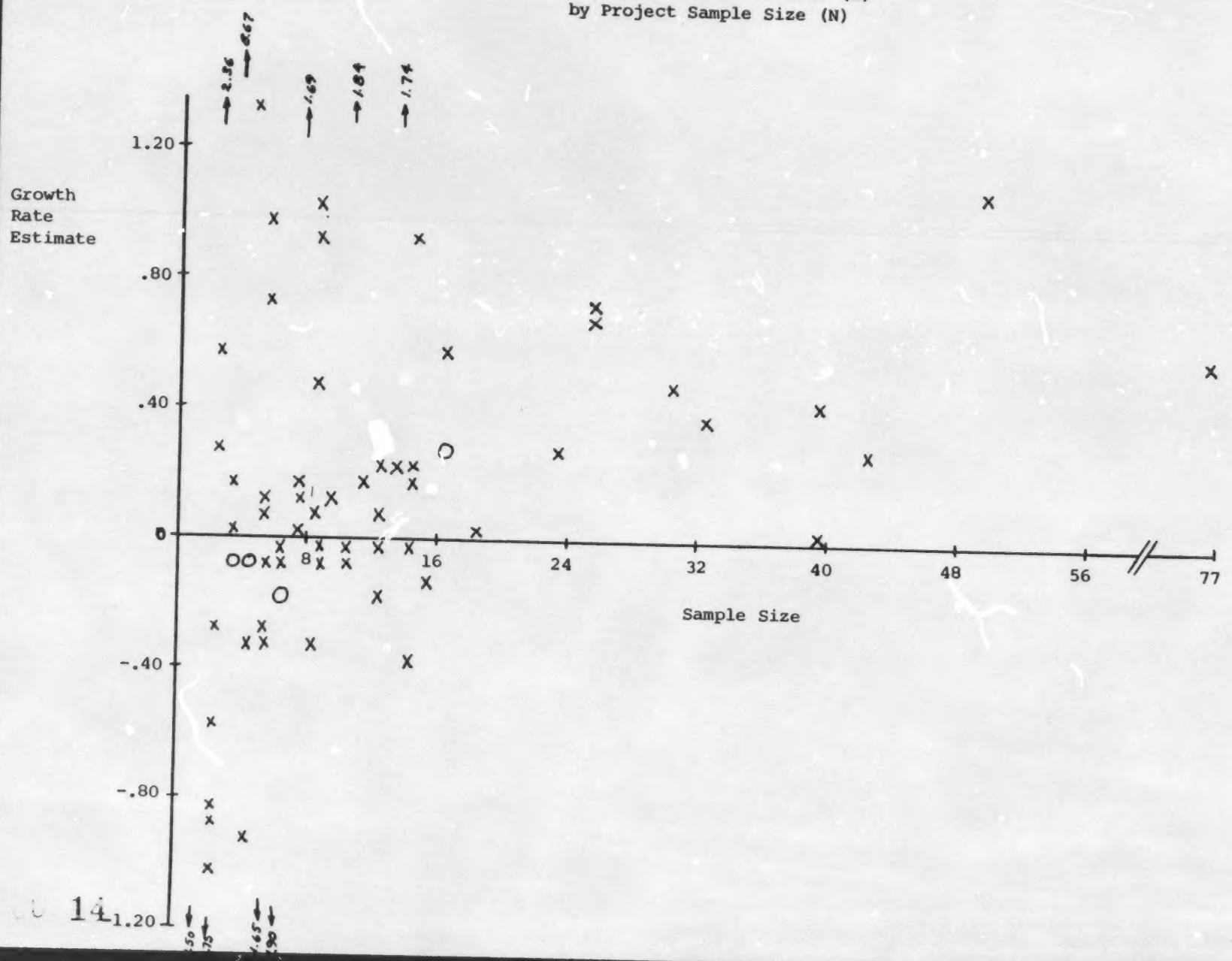


Figure 2

Scatterplot of Growth Rate (b)
by Project Sample Size (N)



the estimates of project impact themselves are the object of the aggregation. The strategy adopted here involved aggregation at an earlier stage, prior to the calculation of project impact. Its objective was to obtain a more stable estimate of growth rate (i.e., no treatment expectation) first, and then proceed to calculate impact estimates for each project on this basis. While this seems an obvious and necessary step, it is not without methodological risk. Pooling (statistically) heterogeneous subgroups of data in performing correlational analyses is known to result in distorted estimates of interrelationship (Sockloff, 1975), and often estimates a relationship which is not of fundamental interest to the researcher. Indeed, Cronbach (1976), in a treatise on the appropriate level (individual or group) of analysis in evaluation studies such as these has noted the frequent misapplication of an individual level (i.e., students) analysis to address a group level (i.e., classroom or school) hypothesis. The interrelationship of interest here--age and cognitive development of young children--is certainly subject to such question. The position taken in this study is that the relationship of interest is at the individual, rather than group, level. Still, the heterogeneity of the groups of data pooled is a consideration to be employed in judging the meaningfulness of analysis and results at the aggregate level.

In all, five methods of aggregation were devised for these data. The justification for any of these methods rests both on conceptual as well as the aforementioned statistical grounds. The conceptual justification alluded to here deals less with the data themselves and more with the content of the various programs. Obviously, a reading program is not a reading program is not a reading program. To the extent that various

programs share similar instructional goals, objectives and activities it seems more conceptually reasonable to merge their data. Thus, before recommending any of the following aggregation methods, one should question their reasonableness on these bases. If the particular method passes this test, statistical issues can be considered. Perhaps a useful analogy is that of selecting a test for use as an evaluation tool in a school program. If a test's content matches the program curriculum well enough, other more quantitative issues can be broached (e.g., test and subtest reliability, inter-level scaling adequacy). If the content validity is not there, however, these other statistical issues become moot.

In describing the five aggregation methods below, the reader may find it easy to get lost in the shuffle. They represent various combinations of data aggregation across grade levels, schools, and districts. In formulating the methods, it seemed obvious that two of a project's characteristics were not aggregable--subject matter and test. That is, it seemed conceptually unreasonable to merge age/pretest data from reading and math programs to estimate a natural growth process. Secondly, the varying characteristics (both content and psychometric) of tests used by the ECE projects easily dissuaded one from pooling data obtained from different tests. Summarily, then, one can read through the following aggregation schemes with the understanding that all resultant groups are from programs of the same subject matter employing the same test.

Method 1 - pool data from the same grade across different schools within a district

Method 2 - pool data from the same school across different grades

Method 3 - pool data across grades and schools within the same district

Method 4 - pool data from the same grade across schools and districts (for the same subject matter and test used, as noted above)

Method 5 - pool data across grades, schools and districts (again, same subject matter and test used)

Of the five methods described, the first three are within-district aggregations, while the last two pool data across different districts. The former may seem more reasonable from the conceptual point of view discussed earlier. The statistical results will be presented for all, however, as they are the focus of this study.

The number of projects and students comprising the aggregates obtained for each of the five methods are shown in Table 3. In many cases, a given pool of data represents an aggregate grouping in more than one method. For example, if a given district has only first grade ECE programs, the merging of all of these data represents an application of method 1 (same grade across schools within a district) and method 3 (across grades and schools within a district). Similarly, if a single project used a particular test that no other project used, its data can be thought of as the aggregate data pool for all five methods. A convention adopted here to deal with the fact that the same collection of projects can represent an aggregate in more than one aggregation method, is to list the collection as an aggregate grouping in the first method it appears, and not repeat it in subsequent aggregates. Thus, the pool of data mentioned in this first example above would appear as an aggregate of method 1 only.

Insert Table 3 about here

Table 3

Composition of Aggregates of
Oregon ECE Project Sample

Aggregate Method	Description	Subject	Pretest (Level)	Grade	# of Proj.	# of Students	Age/Pretest r	Age/Pretest b
1	Across Schools	R	SRA '78 (A)	1	2	7	.29	.86
	in LEA	R	Gates '78 (BR)	1	2	20	.48	.42
	Same Subject,	R	MRT '78 (II)	1	2	35	.26	.89
	Test & Grade	R	Santa Clara '74	1	2	37	.23	.30
		O	Early In '76	K	2	18	.37	1.08
		M	PIAT '70	1	2	20	.48	.21
		R	PIAT '70	1	2	21	.72	.51
		R	Gates '78 (BR)	1	3	43	.35	.50
		R	CTBS '73 (B)	1	2	9	-.01	-.02
		R	MAT-I '78 (Pr.)	1	2	17	-.22	-.20
		R	Gates '78 (BR)	1	5	39	.44	.76
		R	Peabody PVT '59	K	4	109	.21	.37
		R	Peabody PVT '59	1	2	9	-.06	-.23
		R	MAT-S '78 (Pr.)	1	9	65	.18	.15
		M	MAT-S '78 (Pr.)	1	8	46	.01	.01
		R	MAT-S '78 (Pr. 1)	1	3	20	-.20	-.05
2	Across Grades,	O	Boehm	Pre-K, K	2	46	.43	.49
	Within School &	R	Peabody PVT	K,1	2	18	.65	.73
	LEA; Same Subject	R	Peabody PVT	K,1	2	44	.19	.30
	& Test							
3	Schools Within	O	Early In '76	Pre-K, K	3	32	.76	1.70
	LEA; Same Subject	R	Peabody PVT '59	K,1	6	118	.29	.44
	and Test							
4	Across Schools &	R	SRA '78 (A)	1	3	22	-.10	-.22
	LEAs; Same Grade,	R	MRT '76 (I)	1	2	9	-.14	-.16
	Subject & Test	R	Gates '78 (BR)	1	14	133	.39	.50
		R	MRT '76 (II)	1	5	55	.04	.09
		R	CTBS '73 (B)	1	3	12	.71	.78
		M	MAT-S '78 (Pr.)	1	9	62	.19	.17
		O	Early In '76	K	3	37	-.06	-.40
5	Across Grades,	R	MRT '76 (I)	K,1	4	58	.57	1.50
	Schools & LEAs;	O	Early In '76	Pre-K, K	4	51	-.08	-.29
	Same Subject							
	& Test							

Some initial results from correlational analyses of the aggregate level data are encouraging, at least from a statistical estimation standpoint. Of the 31 projects having a negative growth rate estimate (as shown in Figure 1), 26 are members of higher order aggregates. The five which are not each used a test uncommon to any other project. Of these 26 projects, 22 belong to aggregates yielding positive growth rate estimates. Thus, adopting a minimal requirement of a positive growth rate estimate before applying the Value Added model excludes only nine of the 86 projects if the aggregation methods are deemed acceptable.

RESULTS AND DISCUSSION

The calculation of project impact (value added) for each of the projects followed the formulation presented by Bryk & Woods (1980):

$$\text{Value-added} = \text{observed gain} - \text{expected gain}$$

Where:

$$\text{Expected Gain} = \text{estimated growth rate} \times \text{program duration}$$

For example, if the age/pretest regression yielded a growth rate estimate (b) of .80 raw score points per month, and the length of the program was 6 months, the expected gain from pretest to posttest would be .80 x 6 = 4.80 raw score points. If the observed gain (posttest-pretest) was 6.50 raw score points, the value added estimate of project impact would be 6.50 - 4.80 = 1.70, again, in raw score units.

The value-added estimate calculated for many of the projects was not satisfactory, however, due to the instability of many of the growth rate estimates at the project level. To address this concern, the value added calculation was repeated for each project using the growth rate estimates

from each of the higher order aggregates of which it was a member. In other words, the no-treatment expectation for a given project was estimated using the age/pretest data from all projects in the aggregate. Two conditions precluded any of these value added calculations: negative growth rate estimates and posttest level different from pretest level.

The results of all value-added calculations are contained in Appendix B. As indicated above, several projects have more than one value added estimate. For example, project 5 has three--one based upon its own project-level data (est. V.A. = 6.52 raw score points), a second based upon the growth rate estimate from all first grade reading programs in the district (est. V.A. = 9.81 raw score points), and a third based upon the growth rate estimate from all first grade reading projects using the same test in the statewide sample (est. V.A. = 9.31 raw score points). At first glance, the variability of the estimates of impact for a given project is striking. Even discounting those cases in which project-level growth rate estimates are negative, it is apparent that the evaluation of project impact can vary a great deal according to the pool of data (aggregate) in which its growth rate estimate is based. From a policy perspective, it would seem that within-district estimates (i.e., aggregates 1, 2 or 3) would be more acceptable than those obtained data pooled across districts (aggregates 4 and 5). From a more statistical perspective, any method merging data which can be said to be from samples of the same population, is preferable. The specification of that population is, of course, the sensitive issue (e.g., same grade level? same age and achievement range?). Statistical methods are available to test for homogeneity of samples along these criteria, but these steps are viewed as secondary, perhaps confirmatory, steps to assessing the validity of the aggregation on more substantive bases.

The presentation of these value added calculations demonstrates the variability in resultant estimates of project impact likely to be encountered when such aggregation measures are taken. One is left with the question as to which is the "best" estimate of project impact. Perhaps more operationally, from a statewide implementation perpsective: which estimate is to be reported as that representing the effectiveness of the ECE program? In this study, the growth rate estimate used to calculate the value added for each project was selected on the basis of four criteria: sample size (N), positive value of the growth rate estimate (b), the standard error of estimate (s_e) and the "proximity" of the data pool to the actual project in question. Figure 3 is a flowchart representing the decision process. First, the project level growth rate calculation was considered. If it was based on a sample size of 15 or larger, and was positive, it was used to calculate the value added. Fifteen of the 86 projects met these criteria. Secondly, within-district aggregations were considered (aggregation methods 1-3). The same criteria for sample size and positive growth rate estimates were applied to these. If more than one within-district aggregation met these standards, the one with the smallest standard error was selected for subsequent value-added calculation. Thirty-six of the 86 projects arrived at impact estimates in this way. Finally, if neither project-level nor within-district aggregates could supply growth rate estimates which met these standards, extra-LEA aggregates (methods 4 and 5) were considered. Using the same process as in the within-LEA aggregates, growth rate estimates were found for nine more projects. In all, positive growth rate estimates, based upon a sample size of 15 or more at the project or some aggregate level, were found for 60 of the 86

projects in the sample. Of the 26 projects for which value-added impact could not be estimated using these decision rules, 16 were insufficient sample size and did not belong to a higher order aggregate, while 10 had negative growth rate estimates.

Insert Figure 3 about here

The "best estimate" value added calculations are starred (*) in Appendix B for each project. Although 60 projects had growth rate estimates available, four of these used a different test or test level at posttest than at pretest, and are thus not included in the value-added calculations.

In viewing the list of calculations and evaluation results in Appendix B from a centralized reporting perspective, a number of interpretation and reporting issues arise. The remainder of this paper considers some of these and, where possible, offers some empirically-based suggestions as to their resolution.

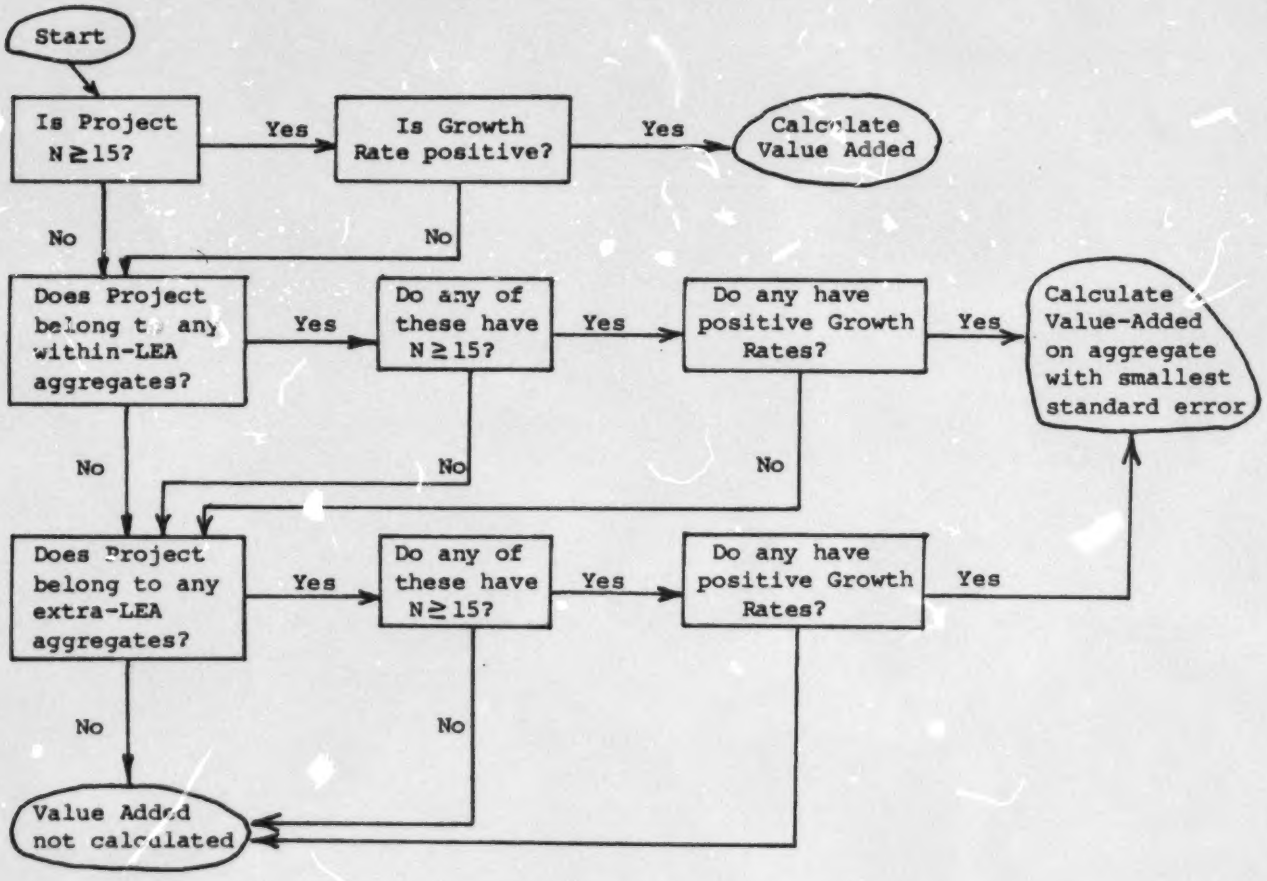
Outcome Metric

The value added calculations in this study were done on raw pretest (and posttest) scores, yielding an outcome in raw score units. While this may be somewhat interpretable at the local level, assuming a high level of familiarity with the instrument, they are not useful at a more central level, and certainly not aggregable. Alternative outcome metrics, suitable for aggregation at a more central level, became a side interest in this study of statewide implementation. Five were considered:

1. Convert the raw score gain to a gain in a norm-referenced score metric (percentiles, NCEs, etc.)

Figure 3

Flowchart of Decision Rule for Selecting growth rate estimate for Value-Added Calculation



2. Express raw score gain as a proportion of the standard deviation of the norm population.

3. Express raw score gain as a proportion of the standard deviation of pretest scores.

4. Express raw score gain as a proportion of the expected gain.

5. Express raw score gain as a proportion of the standard error of estimate.

The first two of these were discarded rather quickly as general reporting strategies. Most of the tests represented in this sample either had no norms available, or what was available was of questionable representativeness (e.g., 600 students from Tennessee and Virginia). The third, expressing gain as a proportion of treatment group pretest standard deviation, was considered promising and useful, but seemed less informative than the next approach, using the standard error of estimate--itself a function of pretest standard deviation. By expressing the gain as a proportion of the standard error of estimate in the growth rate modelling, both the variability of the treatment group at pretest and the strength of the age/pretest relationship are considered:

$$s_e = s_y \sqrt{1 - r_{xy}^2}$$

s_e = standard error of estimate
s_y = standard deviation of pretest scores
r_{xy} = correlation between age and pretest

This could be a misleading index for small sample sizes when the correlation estimate itself is of such high standard error that it often takes on unreasonably high values (as evidenced in the initial correlational results of this study), thereby giving very low standard error of estimate indices. Given the minimal sample sizes required for value added calculations here, however, the standard error of estimate

seemed a useful background for standardization of the raw score calculation of project impact.

$$z_e = \frac{\text{Value Added}}{s_e}$$

Finally, expressing the raw score value added by the gain expected through modelling the growth rate was an idea proposed by Bryk & Woods (1980) in the value-added context, but also suggested by others as an interpretive aid in norm-referenced Title I evaluations (e.g., Stenner, et al, 1977; Gabriel, 1981). The calculation of this index, termed "percent additional growth" (PAG) is as follows:

$$\text{PAG} = \frac{\text{Value Added}}{\text{Expected Gain}}$$

Thus, the impact estimate obtained by the model is expressed as a proportion of the growth expectation, also established by the model.

To illustrate different methods of standardizing the raw score value added estimate, three of these methods were applied to a collection of the projects from this sample. All were first grade reading projects using the Gates MacGinitie Basic Reading Test--one of the only tests represented which had both fall and spring empirical norms at this grade level. This allowed the norm-referenced expression of outcome to be included. The projects selected were from three different districts, and growth rate estimates used were all at the district level (aggregation method 1).

Insert Table 4 about here

Table 4

Three Alternative Standardized Outcome Measures
For Value-Added Estimates

Project*	$\bar{Y}_{Pre}$	$\hat{Y}_{Post}$	$\bar{Y}_{Post}$	V.A.	PAG	Z_e	NCE
4	25.40	28.38	38.00	9.62	323	1.59	16
5	21.94	24.54	34.35	9.81	111	1.28	24
39	19.50	23.00	39.86	16.86	334	2.49	29
40	15.00	18.50	40.37	21.87	966	3.80	39
41	25.00	28.50	43.00	14.50	414	2.20	26
52	36.54	40.11	44.81	4.70	132	.56	11
53	18.33	23.80	38.13	14.33	262	1.70	26
54	29.11	34.20	46.60	12.40	244	1.47	24
55	16.00	21.32	23.14	1.82	34	.22	3
56	27.50	32.67	35.50	2.83	55	.34	5

* All projects are first grade reading, using the Gates '78 (Basic Reading). Growth rate values used were those estimated at district level.

Table 4 contains three standardized outcome measures for these 10 projects--percent additional growth (PAG), standard error units (Z_e), and gains converted to NCEs. All are tied to the value added estimate--PAG derives its growth expectation from the model, Z_e derives its standard error from the age/pretest relationship, and the NCE gain is calculated as the difference between the observed posttest mean ($\bar{Y}_{post}$) and the no-treatment expectation for posttest mean ($\hat{Y}_{post}$). Inspection of these values in Table 4 shows a wide range in interpretation of program effectiveness--percent additional growth ranging from 34% to 966%, for example. The standard error units seem intuitively more reasonable, with project impact estimates ranging from about 1/4 standard error to almost four standard error effects.

Insert Figure 4 about here

Applying this method to the "best" estimates of project impact for all projects, as listed in Appendix B, yields the results shown in Figure 4. Virtually all impact estimates are positive with the distribution showing a positive skew extending to as high as +5 standard error units. The median value is approximately +1.5 standard error units above the no-treatment expectation--again, highly positive but a believable summary of the impact of a collection of ECE programs.

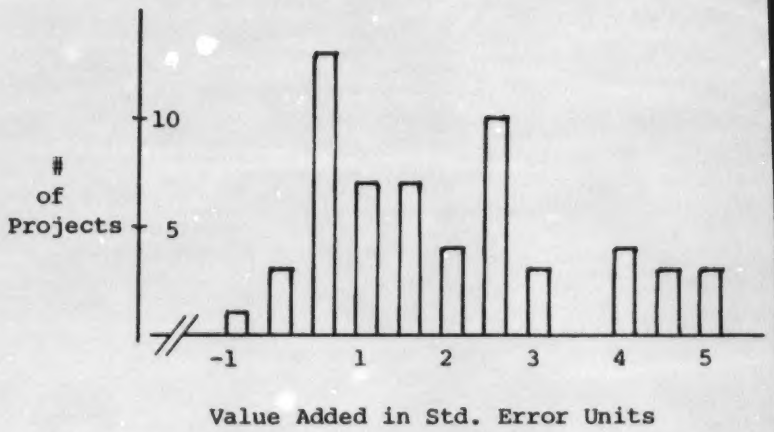
Another Method for Estimating Growth Rate

The seemingly unrealistically high values for PAG and gains expressed in NCE units are less a fault of these indices in themselves than the model's estimate of expected growth. That is, underestimates of growth rate will obviously lead to high values of PAG. One can look at the

Figure 4

Distribution of "Best Estimate"
of Value Added in Standard Error Units (Z_e)

Interval of Z_e Values	f_i
+4.51 to 5.00	3
+4.01 to 4.50	3
3.51 to 4.00	4
3.01 to 3.50	0
2.51 to 3.00	3
2.01 to 2.50	10
1.51 to 2.00	4
1.01 to 1.50	7
.51 to 1.00	7
.01 to .50	13
.00 to -.50	3
-.51 to -1.00	1



model's estimates of growth rate (cf, Appendix B) and see a number of obviously unacceptable values (e.g., .07 raw score units per month). Still, how is one to determine more systematically the validity of these methods?

Another method of arriving at an expected growth rate was attempted on these data. Due to the general ex post facto nature of this study, this is not offered as a pure comparison, but one presented for illustrative purposes to demonstrate the magnitude of the differences resulting from two commonly used evaluation approaches. The validity of the alternative approach is itself a subject of some controversy (e.g., Linn, 1979; Wood, 1980).

The Norm-Referenced evaluation model in TIERS establishes its estimate of growth by assuming that project students would maintain the same percentile/NCE from pretest to posttest given only regular instruction. By selecting projects using a norm-referenced test which has empirical norms at both pretest and posttest dates, the impact estimates obtained from the Value Added model may be compared with those obtained via the TIERS Norm-Referenced model. The projects used in the previous section on outcome metrics are a useful choice. The wide range of PAG values previously noted raised some concern as to the validity of the model's expected growth estimate. Table 5 presents relevant calculations (growth rate estimates (b), expected growth (EG) over the course of the project, additional growth (AG) observed and percent additional growth (PAG)) on these data under both models. In raw score units, the Value Added model's growth expectations range from 2.60 to 5.47 points over the course of the 10 programs (6-7 months in all but one case). The Norm-Referenced model's estimates range from 7 to 16 raw

score points. The growth rates (b) used in the latter were approximated by dividing the number of raw score points needed to maintain the same percentile/NCE pre and post (EG) by the length of the program in months. Note that these are often two or three times larger than the corresponding Value Added estimates. Thus, if one converted the expected posttest derived via the Value Added approach to an NCE, the model would predict a loss in NCE from pretest to posttest. This, again, partly explains why gains derived from the Value Added approach expressed in NCEs were so high (see Table 4).

Insert Table 5 about here

The data chosen for the illustrations in Tables 4 and 5 do not represent the most extreme cases. They do, however, demonstrate a pervasive result in the application of the Value Added model in these data--an apparent underestimate of the no-treatment expectation, and consequent overestimate of project impact. Importantly, this conclusion is reached here not as a truly experimental test of the validity of the model for ECE evaluation. Instead, it is an empirical demonstration of the efficacy of the model in conditions which represent "real world" (i.e., uncontrolled) implementation. Harpring (1982) details our recommendations for more careful implementation of the model--procedures which would undoubtedly provide a fairer test of the model's validity. These recommendations are no strangers to veterans of TIERS or Campbell & Stanley (1963). They require careful considerations as to selection of project participants and instrumentation used to measure program outcomes. The model developers themselves demonstrated much sensitivity

Table 5

Comparison of Results Obtained From
Value Added and Norm-Referenced
Evaluation Models

Project	$\bar{Y}_{pre}$	$\bar{Y}_{post}$	Value-Added				Norm-Referenced			
			b	EG	V.A.	PAG	b	EG	AG	PAG
4	25.40	38.00	.42	2.98	9.62	323	1.97	14	-1.40	-10
5	21.94	34.35	.42	2.60	9.81	111	1.94	12	.41	3
39	19.50	39.86	.50	3.50	16.86	334	1.57	11	9.36	85
40	15.00	40.37	.50	3.50	21.87	966	1.00	7	18.37	265
41	25.00	43.00	.50	3.50	14.50	414	2.00	14	4.00	29
52*	36.54	44.81	.76	3.57	4.70	132	2.13	10	-1.73	-17
53	18.33	38.13	.76	5.47	14.33	262	1.25	9	10.80	120
54	29.11	46.60	.76	5.09	12.40	244	2.39	16	1.49	9
55	16.00	23.14	.76	5.32	1.82	34	1.14	8	-.86	-10
56	27.50	35.50	.76	5.17	2.83	55	2.21	15	-7.00	-46

* Program duration only four months (fall-winter). Interpolated midyear norms used for norm-referenced model calculations

to these issues (e.g., Bryk & Woods, 1980). The current findings support their concerns.

Background Variables

In modelling the natural growth rate of project students, two student characteristics were of interest in this application--sex and preschool experience. Bryk & Woods (1980) recommended calculating growth rate and value added estimates separately for boys and girls, since itis likely that they are at different points in their developmental cycle at the same age. If this is in turn reflected in their pretest scores, their growth rates are likely different as well. The extent of preschool experience among the sample of ECE students was also of interest. A fundamental assumption of the model, the stable universe assumption, dictates that the growth rate estimate determined at pretest reflect natural growth processes in the young children up to that point in time. The presence of prior school experiences would likely affect these processes and invalidate the "natural" growth rate estimate. The intent in this study was to incorporate the presence or absence of prior school experience in the model and at least partly account for its influence on the growth rate estimate.

The inclusion of preschool experience in the model was not a viable alternative in these data due to the high rate of missing data (see Table 1). Even a thorough search of the aggregates yielded a maximum sample size of 18 students with available data on pretest score, age and preschool experience. On these, admittedly meager, data there was no evidence that preschool experience moderated the influence of age in pretest score.

Students' sex was included in the model to determine whether it moderated the growth rate estimate. Multiple regression analyses involving age and sex as predictors rarely added anything to the prediction of pretest score using age alone. Males and females did not differ substantially on their pretest means. Their ages often differed, however, and the regression of pretest on age subsequently was different for males and females. Table 6 contains an illustration from a language development project using the Boehm Test of Basic Concepts in Grades K and 1.

Insert Tables 6 & 7 about here

Table 7 presents a similar calculation for a kindergarten reading project using the Peabody Picture Vocabulary Test. In both of these calculations, the value added estimates of project impact are roughly the same for both sexes. The growth rate estimates, and hence the no-treatment expectations, are very different, however. In Table 4, a K-1 program using the Boehm, the expected growth for boys is twice that for girls. The example in Table 5 shows the opposite trend--the expectation for females is 1-1/2 times that of males. While the raw score value added estimate is not appreciably different for males and females in either case, the difference in growth expectation may lead one to evaluate these outcomes for males and females differently.

Test Characteristics

In viewing the volumes of data and results obtained in this application, the author was continually struck by the influence of the test (and level) used on the believability of the cross-sectional

Table 6

Growth Rate (b) and Value Added Estimates
Obtaines Separately for Males and Females
Boehm Test of Basic Concepts

	Males	Females
N	31	15
Age at Pretest	57.52	55.80
Pretest Mean	30.36	33.73
r	.57	.26
b	.63	.31
Project Length	8.40	8.40
Expected Gain	5.29	2.60
Observed Gain	9.87	6.73
Value Added	4.58	4.13
PAG	87%	159%

Table 7

Growth Rate (b) and Value Added Estimates
Obtained Separately for Males and Females
Peabody Picture Vocabulary Test

	Males	Females
N	47	27
Age	65.94	67.59
Pretest Mean	49.40	47.11
r	.31	.58
b	.47	.69
Project Length	7.00	7.00
Expected Gain	3.29	4.83
Observed Gain	9.78	10.21
Value Added	6.49	5.38
PAG	297%	111%

modelling of growth rate. A few observations from the 86 projects' worth of data are offered here.

There was little indication of floor effects present in pretest data and, relatedly, little evidence of curvilinear age/pretest relationships. The few that were in evidence occurred in group administered and less standardized tests. These instances conform to the well documented cautions in testing young children (Haney & Gelberg, 1980). Ceiling effects on the posttest were in more frequent evidence, however. In most cases, these occurred when projects used the same level at posttest as they did at pretest, even though the test publisher's recommendation was for the next higher level at posttest. Typically, a ceiling effect would tend to depress project impact estimates. In these data, however, limitations imposed by the ceiling effects on the posttest were probably more than compensated for by the low no-treatment expectations previously discussed. Thus, the estimates of project impact via the Value Added approach were not discouragingly low. Still, this is an important issue for careful implementation of the Value Added model. Projects should be encouraged to administer a test level which best matches students' achievement level at the time of testing. If a change in level is thus suggested, a different score metric may be used for the Value Added calculation (e.g., an expanded scale score, if available).

The administration and scoring characteristics of the test also had some influence on results. Some of the tests for young children are group-administered, some are individually administered. While the latter seems obviously preferable for this age group, even these administrations can present difficulties. Tests which are administered until a student misses an item, or two in a row, etc., often showed the lowest

relationships with age. Perhaps test-taking behaviors in young children, independent of their cognitive development, are at their greatest influence in these types of testing situations.

CONCLUDING REMARKS

In summary, this study was designed to present an empirical account of the use of the Value Added Model for evaluating ECE programs. The implementation and reporting conditions were far from ideal, so the results reported here can hardly be considered an experimental test of the validity of the model. These findings are offerred as representing a broadly-based statewide collection of projects attempting to use the model. Although not ideal, the implementation and reporting conditions were real. Other states contemplating use of the model will hopefully benefit from the experiences detailed here.

Some of the major findings of this study include the following:

- Due to a preponderance of small sample sizes, the need for aggregating data across projects to obtain a more stable growth rate estimate was evident. Five methods of aggregation were proposed, and guidelines for selection of the most appropriate growth rate estimate for a given project were suggested.
- The Value Added model provided estimates of expected growth which seemed unacceptably low in most instances.
- Several standardized outcome metrics were suggested. Expressing the raw score value-added as a proportion of the standard error of estimate of the age/pretest relationship was the method of choice.
- Few floor effects at pretest were in evidence. Ceiling effects at posttest were more common.
- Support was found for estimating growth rate separately for males and females in ECE projects. Other background information would likely be useful under more controlled data collection conditions.

- A number of implementation-related characteristics were not known here--student selection procedures, test selection and administration, the nature of the program's activities. Knowledge of these conditions would have contributed to the interpretation of the data. More control of these contributions would have contributed to the validity of the model's use.

- Eighteen different tests were used. The content, administration and scoring of these tests likely affects the efficacy of the Value-Added Model for evaluating project impact.

References

Bryk, A.S. & Weisberg, H.I. Value-added analysis: A dynamic approach to the estimation of treatment effects. Journal of Educational Statistics, 1976, 1, 127-155.

Bryk, A.S.; Strenio, J.F.; & Weisberg, H.I. A method for estimating treatment effects when individuals are growing. Journal of Educational Statistics, 1980, 5, 5-34.

Bryk, A.S. & Woods E. An introduction to the value added model and its use in short term impact assessment. Cambridge, MA: The Huron Institute, 1980.

Campbell, D.T. & Stanley, J.C. Experimental and quasi experimental designs for research. Chicago: Rand McNally, 1963.

Cronbach, L.J. Research on classrooms and schools: formulation of questions, designs and analysis. Stanford Evaluation Consortium Occasional Paper Series. Stanford, California, July 1976.

Gabriel, R.M. Norm-referenced growth expectations: Interpretive cautions for a potentially useful index. Paper presented at the 1981 annual meeting of the California Educational Research Association. San Mateo, California, November 1981.

Haney, W. & Gelberg, W. Assessment in early childhood education. Cambridge, MA: The Huron Institute, 1980.

Harpring, S. Practical problems in the evaluation of early childhood intervention projects. Paper presented at 1982 annual meeting of the American Educational Research Association. New York, New York; March 1982.

Linn, R.L. Validity of inferences based on the proposed Title I evaluation models. Educational Evaluation and Policy Analysis, 1979, 1, 23-32.

Sockloff, A. Behavior of the product moment correlation coefficient when two heterogeneous subgroups are pooled, Educational and Psychological Measurement, 1975, 35, 267-276.

Stenner, A.J.; Hunter, E.L.; Bland, J.D.; & Cooper, M.L. The standardized growth expectation: implications for educational evaluation. Paper presented at the annual meeting of the American Educational Research Association; Toronto, Canada; 1978.

Tallmadge, G.K. & Wood, C.T. User's Guide, Mountain View, CA: RMC Research Corporation, October 1976.

Wood, C.T. The adequacy of the equipercentile assumption in the norm referenced evaluation model. Mountain View, California; RMC Research Corporation, February 1980.

Appendix A

Data Collection Form Used in ECE Evaluation

TITLE I EARLY CHILDHOOD EVALUATION FORM

District Name _____ School Name _____ Grade (check one) ___ Prekindergarten
___ Kindergarten
___ First grade

Subject Matter (List major instructional content areas. Use other side if necessary)

1 _____ 2 _____ 3 _____

Pretest Description _____ (19) _____ _____
name of test edition form level

Posttest Description _____ (19) _____ _____
name of test edition form level

Portion of Test Administered (check one) ___ Subtest only Name of subtest score: _____
___ Total test Name of total test score: _____

Student Information					Pretest Information			Posttest Information		
Student Code Number	Date of Birth (m/d/y)	Sex	Previous School Experience		Date of Admin. (m/d/y)	Raw Score	Standard Score	Date of Admin. (m/d/y)	Raw Score	Standard Score
			Pre-K	K or equiv.						
1			Y N	Y N						
2			Y N	Y N						
3			Y N	Y N						
4			Y N	Y N						
5			Y N	Y N						
6			Y N	Y N						
7			Y N	Y N						
8			Y N	Y N						
9			Y N	Y N						
10			Y N	Y N						
11			Y N	Y N						
12			Y N	Y N						
13			Y N	Y N						
14			Y N	Y N						

Appendix B

Value Added Calculations Using Project-Level and Aggregate Growth Rate Estimates

Proj.	Grade	Subj.	Test (Level)	Aggreg.	N	Pretest Mean	Project Length	b	Expected Gain	S_e	Value Added
1	K	R	SRA '78(A)		0	--	--	--	--	--	--
				2	1						
				3	7						
				5	22						
2	1	R	SRA '78(A)		1	84.00	6.40	--	--		--
				1	7			.86	5.50	15.30	3.00
				4	22			-.22	--	--	--
3	1	R	SRA '78(A)		6	75.33	7.30	.71	5.18	16.92	17.82
				1	7			.86	6.28	15.30	16.72
				4	22			-.22	--	--	--
4	1	R	Gates '78(BR)		5	25.40	7.10	-.36	--		--
				1	20			.48	2.98	6.05	9.62*
				4	133			.50	3.10	7.96	9.31
5	1	R	Gates '78(BR)		15	21.94	6.20	.95	5.89	5.09	6.52*
				1	20			.42	2.60	6.05	9.81
				4	133			.50	3.10	7.96	9.31
6	K	"O"	IRS		77	48.95	8.00	.60	4.80	22.24	34.42*
7	1	R	Gates '78 (BR)		6	22.00	7.00	-.30	--	--	--
				4	133			.50	3.50	7.96	6.50*
8	1	R	MRT '76 (II)		5	27.00	5.30	-.31	--	--	--
				4	55			.09	.48	9.58	-2.50*
9	1	R	MRT '76(II)		9	35.67	8.20	1.01	8.28	11.11	.72
				1	35			.89	7.30	8.94	1.70*
				4	35			.09	.74	9.58	8.26
10	1	R	MRT '76(II)		26	30.31	8.30	.71	5.89	8.17	19.52*
				1	35			.89	7.39	8.94	18.02
				4	55			.09	.75	9.58	24.66
11	1	R	Gates '78(BR)		4	27.40	6.30	.00	.00	6.59	17.35
				4	133			.50	3.15	7.96	14.20*
12	K	R	MRT '76(I)		49	30.51	7.00	1.08	7.56	12.85	18.85*
				5	58			1.50	10.50	13.81	15.91
13	1	R	Santa Clara '74		28	62.76	8.60	.19	1.63	7.08	33.57*
				1	37			.30	2.58	7.55	32.62
14	1	R	Santa Clara '74		9	63.00	8.10	.93	7.53	9.25	24.69
				1	37			.30	2.43	7.55	29.79*

Proj.	Grade	Subj.	Test (Level)	Aggreg.	N	Pretest Mean	Project Length	b	Expected Gain	S_e	Value Added
15	1	R	Gates '78(BR)		10	21.40	7.40	.15	1.11	9.50	1.32
				4	133			.50	3.70	7.96	-1.27*
16	K	"O"	Santa Clara '74		14	73.79	4.00	-.20	--	--	--
17	1	R	MRT '76(I)		8	62.37	12.00	.17	2.04	3.27	18.23
				4	9			-.16	--	--	--
				5	58			1.50	18.00	13.81	2.27*
18	1	M	MATS '78(Pr.)		16	20.50	5.50	.31	1.71	5.88	7.92*
				4	62			.17	.94	4.35	8.69
19	1	R	MRT '65(I)	0							
20	1	R	MRT '65(Pr.)		7	41.71	6.00	-1.90	--	--	--
21	1	R	Woodcock		8	38.62	7.30	1.69	12.34	9.72	29.79
22	1	R	SRA '78(A)		15	60.20	7.30	-.40	--	--	--
				4	22			-.22	--	--	--
23	1	R	Gates '65 (PR A)		3	13.67	7.00	-.82	--	--	--
24	Pre-K	"O"	Early In '76		14	36.57	7.60	1.74	13.22	10.70	1.78
				3	32			1.70	12.92	10.67	2.08*
				5	51			-.29	--	--	--
25	K	"O"	Early In '76		11	58.18	7.60	1.84	13.98	13.05	4.02
				1	18			1.08	8.21	10.94	9.79
				3	32			1.70	12.02	10.67	5.08*
				4	37			-.40	--	--	--
				5	51			-.29	--	--	--
26	1	M	PIAT '70		6	13.00	7.60	.12	.91	.66	5.42
				1	21			.51	3.83	2.93	3.42*
27	1	R	PIAT '70		4	12.75	7.50	-.09	--	--	--
				1	21			.51	3.83	2.93	3.42*
28	K	"O"	Early In '76		7	60.43	7.70	-.03	--	--	--
				1	18			1.08	8.32	10.94	8.82
				3	32			1.70	13.09	10.67	4.05*
				4	37			-.40	--	--	--
				5	51			-.29	--	--	--
29	1	M	PIAT '70		14	14.21	7.70	.24	1.85	2.68	5.65
			5.88	1	20			.21	1.62	2.28	5.88*
30	1	R	PIAT '70		17	14.76	7.70	.56	4.31	3.03	7.12*
				1	21			.512	3.93	2.93	7.50

Proj.	Grade	Subj.	Test (Level)	Aggreg.	N	Pretest Mean	Project Length	b	Expected Gain	S_e	Value Added
31	1	R	CAT '77(10)		5	78.80	7.00	1.37	9.59	23.40	--
32	1	M	CAT '77(10)		4	11.25	7.00	.20	1.40	3.01	--
33	1	R	Gates '78(BR)		12	17.08	7.00	-.03	--	--	--
				4	133			.39	2.73	7.96	--
34	K	"O"	Early In '76		19	3.21	7.00	.01	.07	1.35	6.74*
				4	37			-.40	--	--	--
				5	51			-.29	--	--	--
35	1	R	MRT '76(II)		1	57.00	6.70	--	--	--	--
				4	55			.04	.09	9.58	11.40*
36	1	R	MRT '76(I)		1	39.00	7.30	--	--	--	--
				4	9			-.16	--	--	--
				5	58			1.50	10.95	13.81	19.05*
37	Pre-K	"O"	Boehm		30	28.93	8.40	.48	4.03	8.10	5.81*
				2	46			.49	4.12	7.33	5.70
38	K	"O"	Boehm		16	36.19	8.40	-.15	--	--	--
				2	46			.49	4.12	7.33	2.44*
39	1	R	Gates '78(BR)		26	19.50	7.00	.67	4.69	6.30	15.67*
				1	43			.50	3.50	6.60	16.86
				4	133			.50	3.50	7.96	16.86
40	1	R	Gates '78(BR)		16	15.00	7.00	.34	4.69	6.05	22.99*
				1	43			.50	3.50	6.60	21.87
				4	133			.50	3.50	7.96	21.87
41	1	R	Gates '78(BR)		1	25.00	7.00	--	--	--	--
				1	43			.50	3.50	6.60	28.50*
				4	133			.50	3.50	7.96	28.50
42	1	R	MAT '70(Primer)		8	22.86	3.00	-.06	--	--	--
43	1	R	CAT '77 (II)		3	21.33	6.10	.30	1.83	.91	18.84
44	1	R	MRT '76 (II)		14	40.21	8.00	.19	1.52	5.59	--
				4	55			.09	.72	9.58	--
45	1	M	MATS '78(Pre-Pr)		39	18.97	7.00	.02	.14	2.82	10.39*
46	1	R	MATS '78(Pre-Pr)		39	45.95	7.00	.41	2.87	8.19	--
47	1	R	CTBS '73(B)		3	29.33	5.10	2.36	12.04	.97	16.96

Proj.	Grade	Subj.	Test (Level)	Aggreg.	N	Pretest Mean	Project Length	b	Expected Gain	S_e	Value Added
48	1	R	CTBS '73(B)		9	17.56	7.00	-.02	--	--	--
				1	9			-.02	--	--	-
				4	12			.78	5.46	5.86	--
49	1	R	CRBS '73(B)		0						
50	1	R	MATI '78(Primer)		14	11.07	6.10	-.02	--	--	--
				1	17			-.20	--	--	
51	1	R	MATI '78(Primer)		3	8.51	6.30	-.89	--	--	--
				1	17			-.20	--	--	--
52	1	R	Gates '78(BR)		13	36.54	4.70	.21	.99	6.55	7.28
				1	39			.76	3.57	8.41	4.70*
				4	133			.50	2.35	7.96	5.92
53	1	R	Gates '78(BR)		12	18.33	7.20	.06	.43	3.84	19.37
				1	39			.76	5.47	8.41	14.33*
				4	133			.50	3.60	7.96	6.20
54	1	R	Gates '78(BR)		9	29.11	6.70	.42	2.81	3.51	14.68
				1	39			.76	5.09	8.41	12.40*
				4	133			.50	3.35	7.96	14.14
55	1	R	Gates '78(BR)		3	16.00	7.00	.60	4.20	5.66	-.70
				1	39			.76	5.32	8.41	1.82*
				4	133			.50	3.50	7.96	.00
56	1	R	Gates '78(BR)		2	27.50	6.80	-1.50	--	--	--
				1	39			.76	5.17	8.41	2.83*
				4	133			.50	3.40	7.96	4.60
57	K	R	PPVT '59		32	45.47	7.00	.19	1.33	7.18	12.79*
				1	109			.37	2.59	8.71	11.53
				3	118			.44	3.08	8.65	11.14
58	K	R	PPVT '59		11	48.27	7.00	1.30	9.10	7.28	-.46
				1	109			.37	2.59	8.71	6.05
				2	18			.73	5.11	7.14	3.53*
				3	118			.44	3.08	8.65	-.36
60	K	R	PPVT '59		7	57.57	7.00	-1.64	--	--	--
				1	109			.37	2.59	8.71	2.92
				2	44			.30	2.10	9.02	3.41
				3	118			.44	3.08	8.65	2.43
61	1	R	PPVT '59		2	58.00	7.00	8.67	60.69	--	-54.69
				1	9			-.23	--	--	--
				2	44			.30	2.10	9.02	3.90
				3	118			.44	3.08	8.65	2.92*

Proj.	Grade	Subj.	Test (Level)	Aggreg.	N	Pretest Mean	Project Length	b	Expected Gain	S_e	Value Added
62	K	R	PPVT '59		24	50.21	7.00	.33	2.31	7.74	5.00*
				1	109			.37	2.59	8.71	4.72
				3	118			.44	3.08	8.65	4.23
63	1	R	MATS '78(Primer)		3	19.33	7.30	-.57	--	--	--
				1	65			.15	1.10	4.94	10.24*
64	1	R	MATS '78(Primer)		12	8.54	7.30	.22	1.61	1.43	11.77
				1	65			.15	1.10	4.94	12.28*
65	1	M	MATS '78(Primer)		6	15.83	7.30	.98	7.15	1.15	.69
				1	46			.01	.07	3.43	7.77*
				4	62			.17	1.24	4.35	6.60
66	1	R	MATS '78(Primer)		11	11.92	7.30	.20	1.46	4.47	19.87
				1	65			.15	1.10	4.94	20.23*
67	1	M	MATS '78(Primer)		3	15.75	7.30	-.30	--	--	--
				1	46			.01	.07	3.43	7.18*
				4	62			.17	1.24	4.35	6.01
68	1	R	MATS '78(Primer)		3	7.20	7.30	--	--	--	--
				1	65			.15	1.10	4.94	19.50*
69	1	M	MATS '78(Primer)		3	13.33	7.30	-1.75	--	--	--
				1	46			.01	.07	3.43	14.93*
				4	62			.17	1.24	4.35	13.76
70	1	R	MATS '78(Primer)		9	18.40	7.30	.06	.44	4.98	9.66*
				1	65			.15	1.10	4.94	8.60
71	1	R	MATS '78		7	17.29	7.30	-.16	--	--	--
				1	65			.15	1.10	4.94	2.90*
72	1	M	MATS '78(Primer)		5	18.60	7.30	-.40	--	--	--
				1	46			.01	.07	3.43	10.13*
73	K	R	Boehm	--	--	--	7.30	--	--	--	--
74	1	R	MATS '78(Primer)		7	18.29	7.30	-.71	--	--	--
				1	65			.15	1.10	4.94	-2.82*
75	1	M	MATS '78(Primer)		6	18.33	7.30	-.09	--	--	--
				1	46			.01	.07	3.43	9.77*
				4	62			.17	1.24	4.35	8.60
76	1	M	MATS '78(P2)		5	14.60	7.30	-.90	--	--	--

Proj.	Grade	Subj.	Test (Level)	Aggreg.	N	Pretest Mean	Project Length	b	Expected Gain	S_e	Value Added
77	1	R	MATS '78(PR 1)		6	6.67	7.30	0.36	--	--	--
				1	20			-.20	--	--	--
78	1	R	MATS '78(PR 1)		8	6.13	7.30	.12	.88	1.74	24.99
				1	20			-.05	--	--	--
79	1	M	MATS '78(Primer)		3	15.00	7.30	-.98	--	--	--
				1	46			.01	.07	3.43	15.26*
				4	62			.17	1.24	4.35	14.09
80	1	R	MATS '78(Primer)		9	13.22	7.30	-.07	--	--	--
				1	65			.15	1.10	4.94	6.24*
81	1	M	MATS '78(Primer)		11	17.00	7.30	-.07	--	--	--
				1	46			.01	.07	3.43	15.57*
				4	62			.127	1.24	4.35	14.40
82	1	R	MATS '78 (PR 1)		6	7.86	7.30	.06	.44	1.63	17.70
				1	20			0.05	--	--	--
83	1	M	MATS '78 (PR 1)		8	15.13	7.30	.01	.07	4.60	11.42
84	K	R	MRT '76 (I)		0						
				5	58			1.50	--	--	--
85	1	R	MATS '78(Primer)		4	13.50	7.30	-.07	--	--	--
				1	65			.15	1.10	4.94	12.07*
86	1	M	MATS '78(Primer)		9	19.30	7.30	-.33	--	--	--
				1	46			.01	.07	3.43	8.43*
				4	62			.17	1.24	4.35	7.26

*Value-added estimate of choice